

Requesting, Reusing, and Analyzing Data on the NF Data Portal

2024 Global NF Conference; June 21st, 2024

Speakers:

Robert Allaway, PhD | Principal Scientist, Sage Bionetworks

Christina Conrad, PhD | Biomedical Data Manager, Sage Bionetworks

Jineta Banerjee, PhD | Principal Scientist, Sage Bionetworks

Anh Nguyet Vu, BS | Senior Bioinformatics Engineer, Sage Bionetworks

Agenda:

1. Introduction
2. Locating Data on the NF Data Portal
3. Interacting with Data from the NF Data Portal on cBioPortal
4. Demonstration: Using R and Cavatica
5. Work Session: NF Team Assistance with Data Access
6. Closing Survey

Introduction

Who makes it, who uses it?

Sage Bionetworks with support from:



29K File Downloads & 409 unique users in last 12 months

What's on it?

248 studies & growing
drug screening, imaging, genomic variants, gene
expression data & more

Where can I find it?

nf.synapse.org

Why was it created?

The portal was created to accelerate NF research by providing a centralized, accessible platform for data sharing and collaboration.

This mission is to enhance the understanding of NF, facilitate the discovery of new treatments, and improve patient outcomes.

How do I get started?

To get started, visit nf.synapse.org and create an account. Explore the available datasets, utilize the analytical tools, and participate in the research community to advance your NF studies.



Repository of animal models, cell lines, antibodies, genetic reagents, biobanks

nfibromatosis re

l to help openly explore and sha
can join the NF Open Science In

ewsletter

PORTA

Initiatives

Studies

Datasets

Files

Publications

Tools

Hackathon Projects

Documentation on how
to navigate portal and
steps for
uploading/downloading
data

Initiatives

24

Studies

248

Publications

202

Tools

1,147

Initiatives are funder-organized programs, groups, consortia, cohorts, or awards, usually focused on a specific research area in neurofibromatosis.

Studies are hypothesis-driven projects with the goal of uncovering new knowledge about neurofibromatosis type 1, type 2, or schwannomatosis.

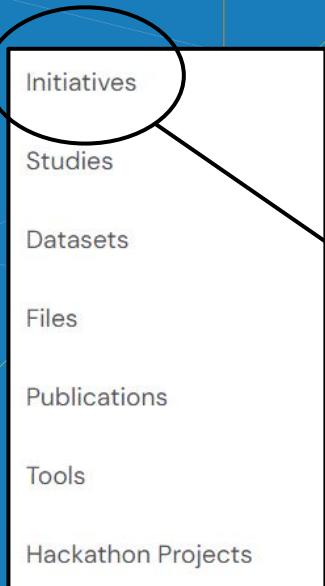
Data are collected from human samples, animal models, and cell lines from a variety of assays.

Publications are pre-prints and peer-reviewed articles produced by NF Data Portal studies.

Find neurofibromatosis research tools: animal models, cell lines, genetic reagents, antibodies, and biobanks!



How to navigate



Initiatives often focus on a type of research and might be a good way to find projects related to a topic of interest



INITIATIVE

Biology and Therapeutic Development for Cutaneous Neurofibromas

Biology and Therapeutic Development for Cutaneous Neurofibroma (BTD cNF) initiative aims to accelerate the next era of discovery and therapeutics to prevent and treat cNF patients with neurofibromatosis type 1 (NFI). The focus areas of this initiative are i) understand the cNF development process, ii) identify and validate therapeutic candidates for cNF, iii) identify and validate non-invasive approaches [...Show More](#)



INITIATIVE

Brain Tumor Initiative

The Brain Tumor Initiative (BTI), funded by Gilbert Family Foundation focuses on identifying and developing therapeutic strategies to effectively treat neurofibromatosis type 1 (NFI)-associated transformed gliomas (non-optic pathway gliomas). BTI is conducting clinical studies to provide critical characterization of NFI high grade glioma and identify the most promising treatments for patients available [...Show More](#)

[Visit Website](#)



Sage Bionetworks

Initiatives

Studies

Datasets

Files

Publications

Tools

Hackathon Projects

Data Status

All



107

Data Pending

69

Under Embargo

46

Available

23

Data Not Expected

2

Partially Available

Show all (6)

Manifestation

All



115

Not Assigned

48

Plexiform Neurofibroma

33

MPNST

29

Cutaneous Neurofibroma

13

Schwannoma

Show all (21)

Disease Focus

All



218

Neurofibromatosis type 1

17

Neurofibromatosis type 2

8

Schwannomatosis

2

Not Assigned

2

Multiple

Show all (6)

Data Type

All



158

Not Assigned

42

geneExpression

39

image

28

genomicVariants

18

drugScreen

Show all (48)

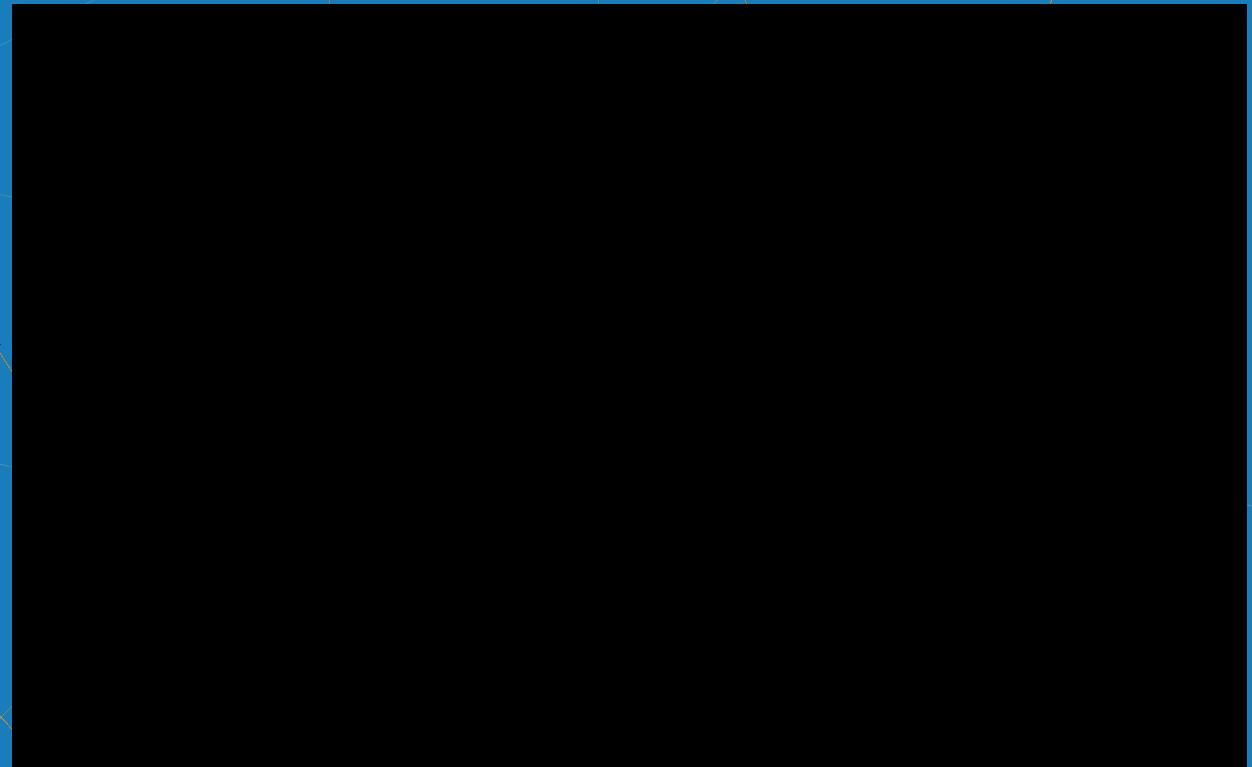
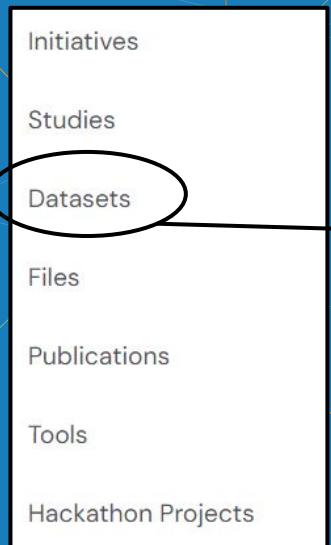
Selecting "**Available**" filters data that is accessible by Synapse users. **Please note:** there may be extra restrictions you need to complete. (Stay tuned)

You can explore what is currently ongoing, even if there's no data. Or, if the data are under embargo, we encourage you to reach out to the study PI to see if they can start a collaboration prior to public data release.

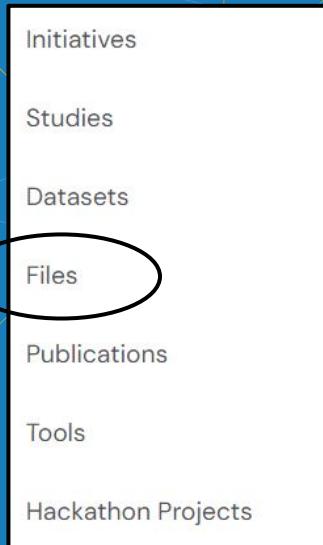


Sage Bionetworks

Selecting **Datasets** is a good way to find analysis-ready processed data on the NF Data Portal



Selecting **Files** allows you to do a broader query across the portal and narrow down the data of interest by assay, data type, file type, etc.



Sign in

NF DATA PORTAL

A home for Neurofibromatosis research resources

The NF Data Portal was created to help openly explore and share NF datasets, analysis tools, resources, and publications related to neurofibromatosis and schwannomatosis. Anyone can join the NF Open Science Initiative (NF-OSI) to contribute.

Subscribe to the NF-OSI Newsletter

PORTAL PROGRAMS AND GOALS

Initiatives	Studies	Files	Publications	Tools
24	249	67,507	202	147

Initiatives are funder-organized programs, groups, consortia, cohorts, or awards, usually focused on a specific research area in neurofibromatosis.

Studies are hypothesis-driven projects with the goal of uncovering new knowledge about neurofibromatosis type 1, type 2, or schwannomatosis.

Data are collected from human samples, animal models, and cell lines from a variety of assays.

Publications are pre-prints and peer-reviewed articles produced by NF Data Portal studies.

Find neurofibromatosis research tools: animal models, cell lines, genetic reagents, antibodies, and biobanks!

Explore

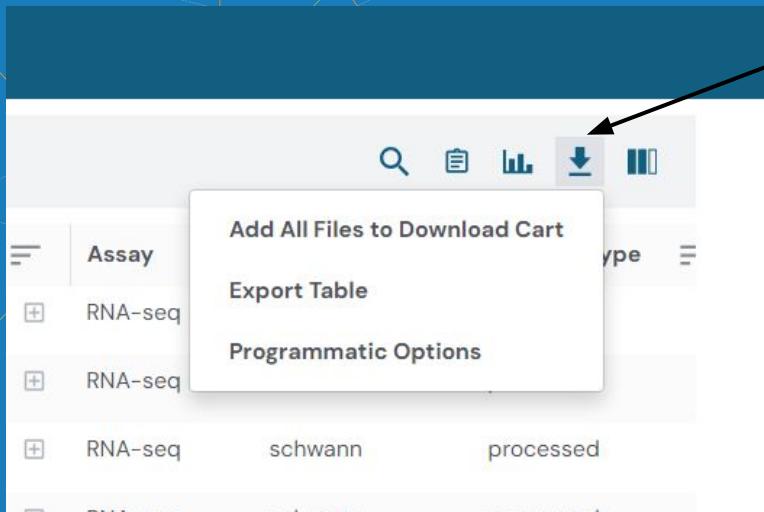
Explore

Explore

Explore

Explore

Accessing Metadata



Export metadata table
to your computer

Control the metadata visible
on the portal by selecting
Show/Hide Columns

A screenshot of a 'Show/Hide Columns' panel on the right side of the portal. It lists various metadata columns with checkboxes: Name (checked), Resource Type (checked), Assay (unchecked), Data Type (checked), Diagnosis (checked), Tumor Type (checked), Species (checked), Individual ID (unchecked), Specimen ID (unchecked), File Format (unchecked), Data Subtype (unchecked), NF1 Genotype (unchecked), NF2 Genotype (unchecked), and On Synapse (unchecked). An arrow points from the 'Export Table' button in the main interface to the 'Show/Hide Columns' panel.



How do I know whether I have access to the data?

Green unlock = you have access!

			File ID	
			<u>syn30058767</u>	
			<u>syn30058789</u>	
			<u>syn30058815</u>	

Golden/Yellow Lock = you need to complete an access restriction or the data is not shared

			File ID	
			<u>syn11887911</u>	
			<u>syn11887914</u>	
			<u>syn11887916</u>	



Encountering a dataset where you need to complete an access restriction

To access the data you will need to complete the following steps:

Step 1: Register for a Synapse account here

Step 2: Request access to Data by clicking the Request Access button below and **submitting your Intended Data Use statement (1-3 paragraphs in English) to be posted publicly here**

Step 3: Review the Terms of Use below and click "Accept Terms of Use"

The screenshot shows the NF DATA PORTAL interface. At the top, there is a navigation bar with links: Browse Tools, Explore, Organizations, About, Contribute Data, News, Help, and a user profile icon. Below the navigation bar, the main content area has a dark teal header with the word "DATASET" and the title "Processed RNA-seq STAR Salmon outputs". Underneath the title, there is a summary table with the following data:

STUDY NAME	The Johns Hopkins NFI Biospecimen Repository
SIZE	712.25 MB
DISEASE FOCUS	Neurofibromatosis type 1
Show More ▾	

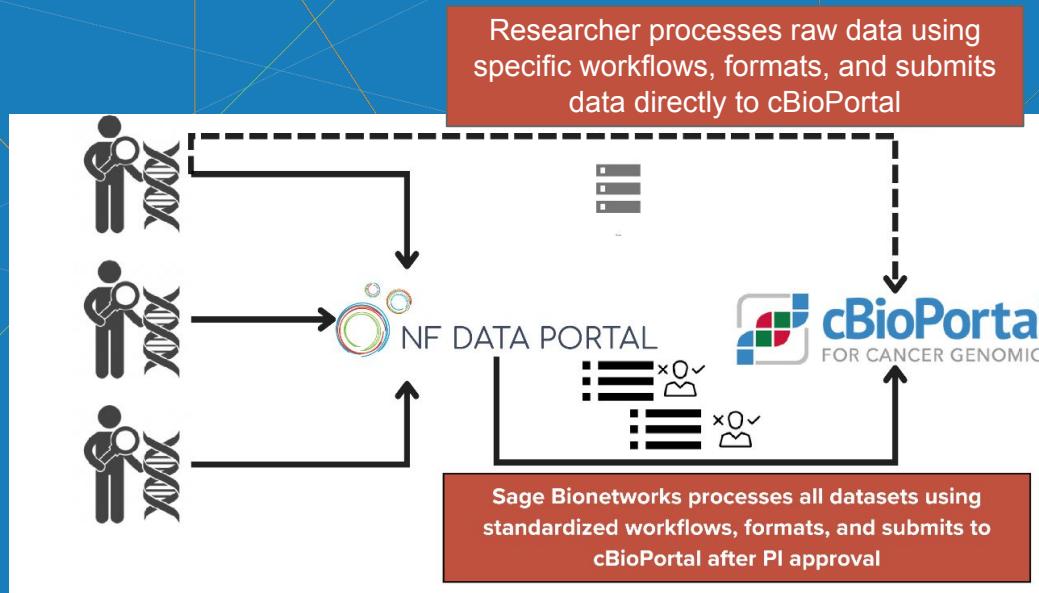
At the bottom of the page, there is a search and filter interface with various buttons and dropdown menus. One of the dropdown menus is labeled "Tumor Type" and includes an option for "Plexiform Neurofibroma".

Can we use data from the NF Data Portal for the following?

User Role	User Story
NF researcher	I want to compare my own data with other studies on the NF Data Portal or cBioPortal.
PI or postdoc	I want to quickly generate a visualization for a grant proposal using publicly available samples/data.
Bench scientist	I cannot work with raw data that is available on NF Data Portal. I want a point-and-click interface to explore the data .
Clinician scientist	I am very familiar with cBioPortal, I want to explore data from the NF Data Portal on cBioportal .



Visualizing NF Data Portal data on cBioPortal



Visualize and analyze NF Portal data alongside other cancer studies

What if I want to access and work with raw data files?

User Role	User Story
NF researcher	I want to compare my own data with other studies on the NF Data Portal
Computational Scientist	I need access to raw data files to run custom analyses and develop new algorithms.





Initializing "my-first-analysis"

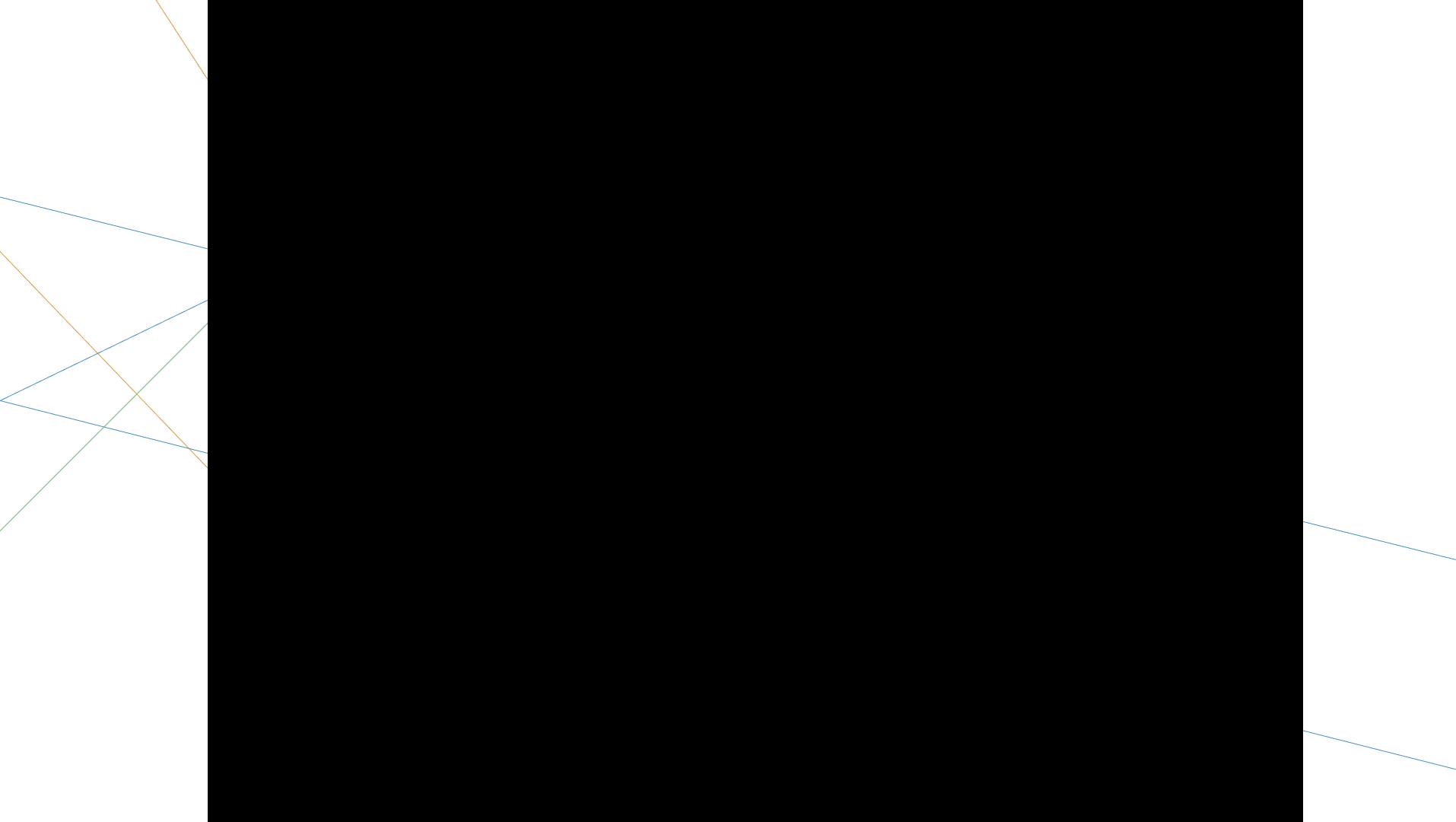
We're almost there. Your analysis execution environment is being set up and initialized.



What if I want to use my local machine to download data?

- You may find advantages to using your local machine (as opposed to a cloud environment like Cavatica to download data)
 - Some advances may include
 - Speed and Performance
 - Full utilization of local hardware capabilities.
 - No additional costs associated with data downloads.
 - Enhanced control over data security.
 - Offline Capability

Next, we will show a tutorial of how to use R on your local machine to download data from the NF data portal



And now a sneak peek of the future...



Sage Bionetworks

Join a pilot project integrating the NF Data Portal with the Pluto.bio platform

Objective: Evaluate Pluto.bio's utility for analyzing and visualizing NF datasets, focusing on low grade glioma (LGG).

Background:

- Need for intuitive, no-code data analysis tools.
- Conversations with the scientific community highlighted this need.

Pilot Details:

- Duration: Roughly 3 months, starting late June/early July.
- Participants: Selected for their focus on LGG, genomic data generation, or bioinformatics experience (both limited and advanced).

Participant Involvement:

- Provided with Pluto.bio accounts and access to LGG datasets.
- Onboarding session at the beginning (multiple session times available).
- 8 weeks of platform access with technical support.
- Interrogate datasets, explore data, ask meaningful scientific questions, and gain new insights.
- At the end of the pilot, feedback on the platform and next steps will be solicited.

Next Steps:

- If interested please submit your interest in the survey at the end of the session. Or email Irene Morganstern (CTF) imorganstern@ctf.org

Questions before working session?

Working Session Suggestions

Option 1 (no-code)

Explore the JH NF1 Biospecimen Repository dataset on cBioPortal.

Learning outcomes:

- Dexterity with one of the main portal entry points, Datasets
- Mastery of facets for filtering and selection of datasets
- Understand linkage of NF data to cBioPortal
- Practiced with analysis of NF data in cBioPortal

Let us know if you need help by flipping your card



Help!



Good



Sage Bionetworks

Working Session Suggestions

Option 2

Search the NF data portal and find data of interest and send the data to Cavatica (for those who've signed up)

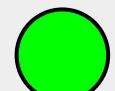
Learning outcomes:

- Mastery of facets for filtering and selection of data files
- Dexterity with one of two qualifying data entry points for Cavatica (global file browser or study-scoped file browser)
- Proficiency with requirements and flow for sending data to Cavatica
- Understand metadata imported with requested assets, how to navigate imported assets in Cavatica
- Experienced with running an R analysis with NF data in Cavatica Data Studio

Let us know if you need help by flipping your card



Help!



Good



Sage Bionetworks

Working Session Suggestions

Option 3

Search the NF data portal and find data to download to your local computer for analysis or visualization

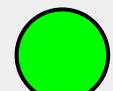
Learning outcomes:

- Mastery of facets for filtering and selection of data files
- Practiced with downloading data to local machine through UI or programmatically
- Understand portal dataset connection <-> Synapse project dataset
- Familiarity with typical Synapse project for additional data navigation
- Ability to install and use synapser (R client for Synapse)
- Briefed on benefits of making a dataset and how to make a dataset
- Experienced with running an R analysis locally with NF data

Let us know if you need help by flipping your card



Help!



Good



Sage Bionetworks

Option 1 (no-code)

Explore the JHU Biobank dataset on cBioPortal.

Learning outcomes:

- Dexterity with one of the main portal entry points, Datasets
- Mastery of facets for filtering and selection of datasets
- Understand linkage of NF data to cBioPortal
- Practiced with analysis of NF data in cBioPortal

Option 3

Search the NF data portal and find data to download to your local computer for analysis or visualization

Learning outcomes:

- Mastery of facets for filtering and selection of data files
- Practiced with downloading data to local machine through UI or programmatically
- Understand portal dataset connection <-> Synapse project dataset
- Familiarity with typical Synapse project for additional data navigation
- Ability to install and use synapser (R client for Synapse)
- Briefed on benefits of making a dataset and how to make a dataset
- Experienced with running an R analysis locally with NF data

Option 2

Search the NF data portal and find data of interest and send the data to Cavatica (for those who've signed up)

Learning outcomes:

- Mastery of facets for filtering and selection of data files
- Dexterity with one of two qualifying data entry points for Cavatica (global file browser or study-scoped file browser)
- Proficiency with requirements and flow for sending data to Cavatica
- Understand metadata imported with requested assets, how to navigate imported assets in Cavatica
- Experienced with running an R analysis with NF data in Cavatica Data Studio

Let us know if you need help by flipping your card



Help!



Good



Sage Bionetworks

Please take our post-workshop survey!
<https://tally.so/r/w4k8Od>



Need NF Data Portal help?

Check out our documentation at
help.nf.synapse.org

Email us at
nf-osi@sagebionetworks.org

Schedule a video chat consult!
<https://calendly.com/d/3ns-7dq-zks/nf-osi-office-hours>



Sage Bionetworks

Navigating from NF Data Portal to cBioPortal

The screenshot shows the NF Data Portal homepage within a web browser. The URL in the address bar is <https://staging.nf.synapse.org>. The page features a dark blue header with the "NF DATA PORTAL" logo and a network graph icon. Below the header, a main banner reads "A home for Neurofibromatosis research resources". It includes a brief description of the portal's purpose, a "Subscribe to the NF-OSI Newsletter" button, and a "PORTAL PROGRAMS AND GOALS" section. This section contains five cards: "Initiatives" (24), "Studies" (248), "Files" (38,070), "Publications" (202), and "Tools" (1,147). Each card has a detailed description below it. To the right of the main content area is a vertical sidebar with various icons and a dropdown menu labeled "C".

PORTAL PROGRAMS AND GOALS

- Initiatives** 24
- Studies** 248
- Files** 38,070
- Publications** 202
- Tools** 1,147

Initiatives are funder-organized programs, groups, consortia, cohorts, or awards, usually focused on a specific research area in neurofibromatosis.

Studies are hypothesis-driven projects with the goal of uncovering new knowledge about neurofibromatosis type 1, type 2, or schwannomatosis.

Data are collected from human samples, animal models, and cell lines from a variety of assays.

Publications are pre-prints and peer-reviewed articles produced by NF Data Portal studies.

Find neurofibromatosis research tools: animal models, cell lines, genetic reagents, antibodies, and biobanks!



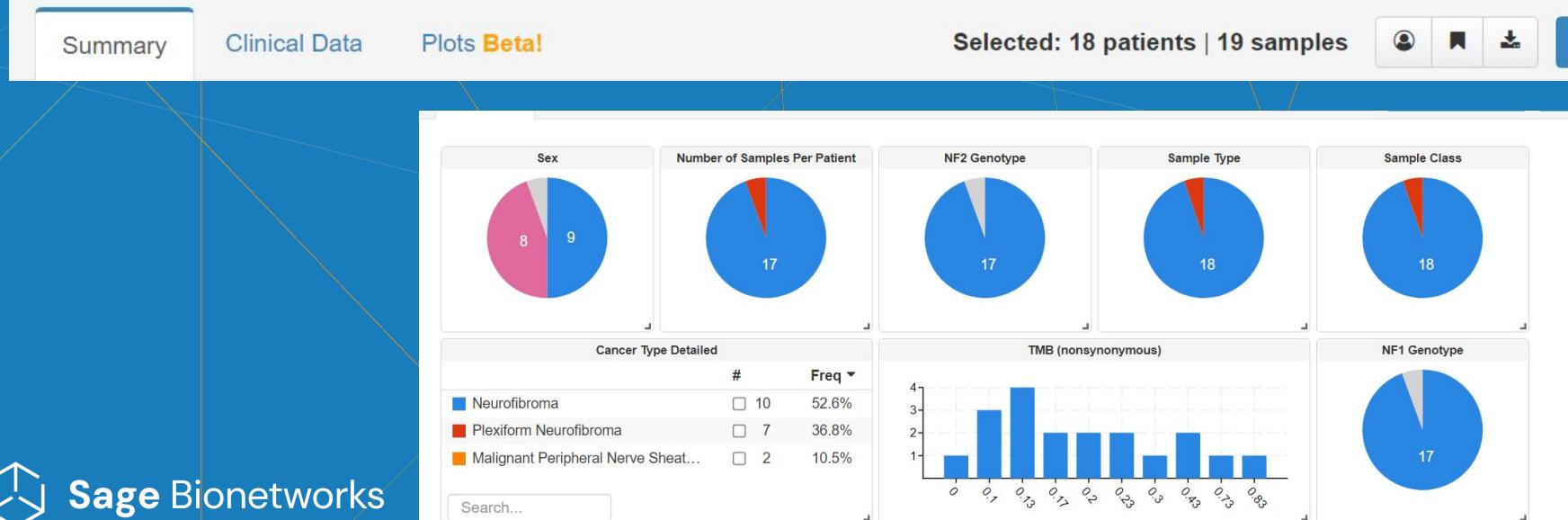
Sage Bionetworks

Visualize the project data on cBioPortal

Nerve Sheath Tumors (Johns Hopkins, Sci Data 2020) [Download](#)

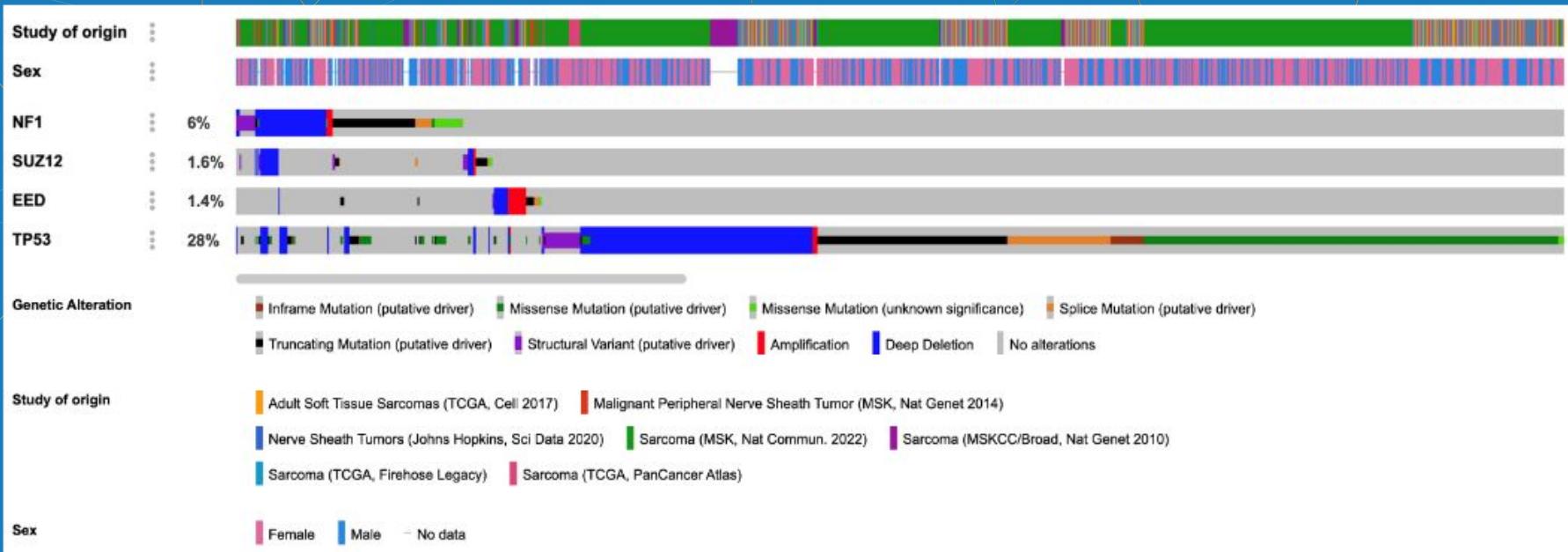
Whole-exome sequencing of 18 patients with matched nerve sheath tumor-normal samples. The raw data is contributed by Johns Hopkins University researchers (Pollard et al. Sci Data 2020) funded by the Neurofibromatosis Therapeutic Acceleration Program (NTAP). The reprocessing of the raw data is managed by the NF Open Science Initiative ([NF Data Portal](#)). The original raw data was reprocessed with updated genome build and filtered to remove low confidence variant calls, and any common variants that had a gnomAD allele frequency (for any subpopulation) greater than or equal to 0.0005. [PubMed](#)

Click

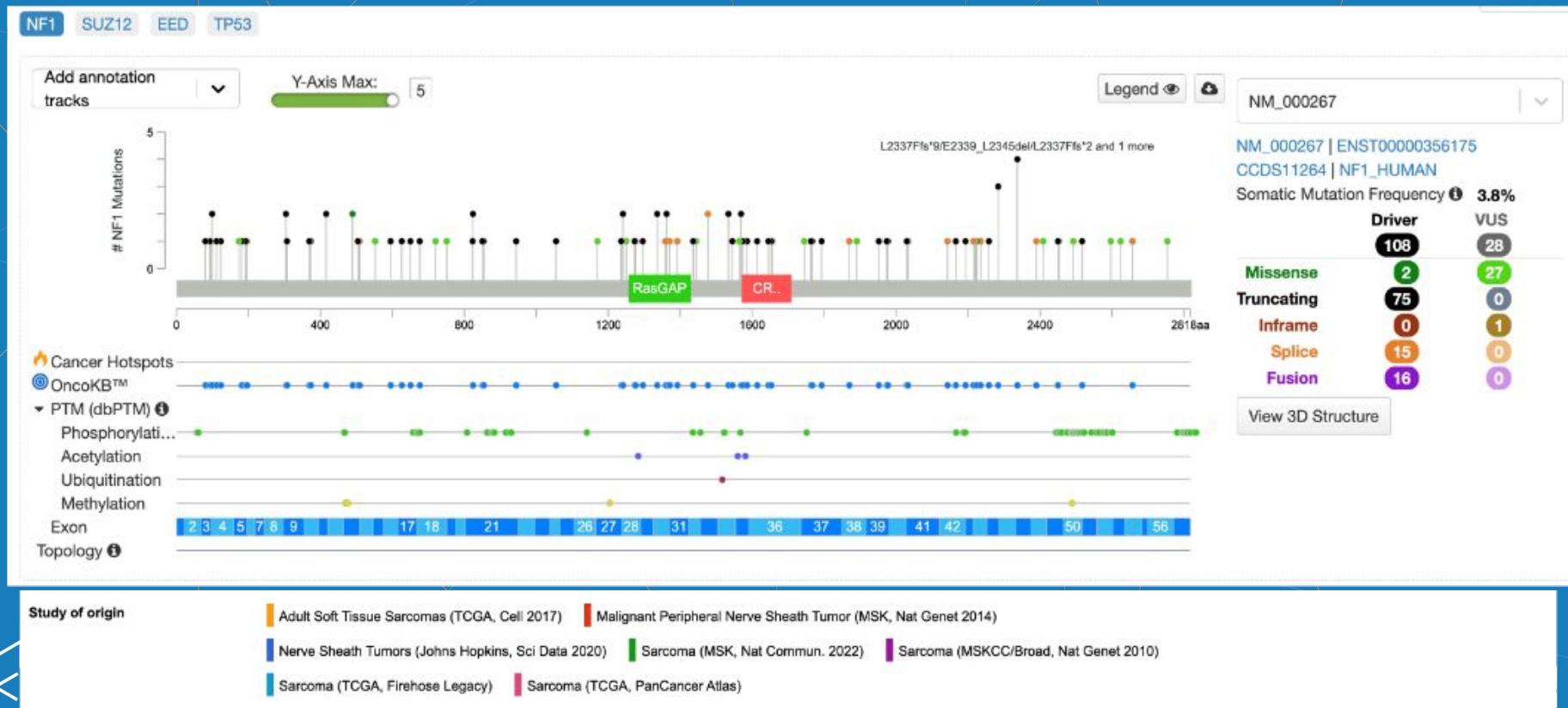


Sage Bionetworks

Analyze the Portal data alongside other cancer studies



Analyze the Portal data alongside other cancer studies



Data Curation using Large Language Models

2024 Global NF Conference; June 21st, 2024

Speakers:

Robert Allaway, PhD | Principal Scientist, Sage Bionetworks
Christina Conrad, PhD | Biomedical Data Manager, Sage Bionetworks
Jineta Banerjee, PhD | Principal Scientist, Sage Bionetworks

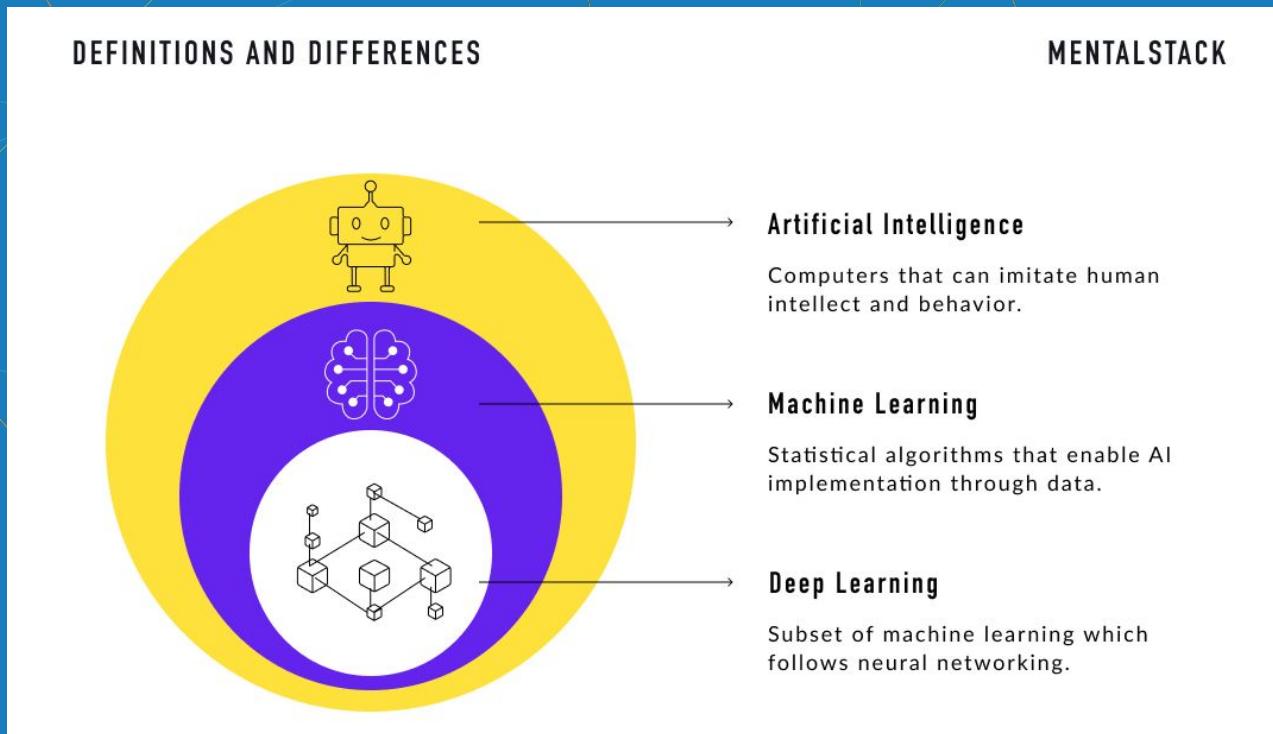


Sage Bionetworks

Agenda:

1. An AI primer
2. Examples of how we use LLMs
 - a. Plain Language Summaries
 - b. NF Tools Central
3. Work Session: Curation exercises with roving support

What are AI, ML, deep learning?



Types of Generative AI

Text Generation:

Creating coherent and contextually relevant text.

Example Models: GPT-4o, Llama 3, Claude Opus

Image Generation:

Creating new images from scratch or modifying existing images.

Example Models: DALL-E, Generative Adversarial Networks

Music Generation:

Composing original music or modifying existing tracks.

Example Models: ElevenLabs Music

Video Generation:

Creating new video content or animating existing images.

Examples: OpenAI Sora

Voice and Speech Generation:

Creating human-like speech and voice recordings.

Example Models: Meta Voicebox, ElevenLabs Multilingual v2

Code Generation:

Writing and generating programming code.

Example Models: Codex, GitHub Copilot.



Video Generation, about a year and a half ago



Source: "chaindrop" Reddit



Sage Bionetworks

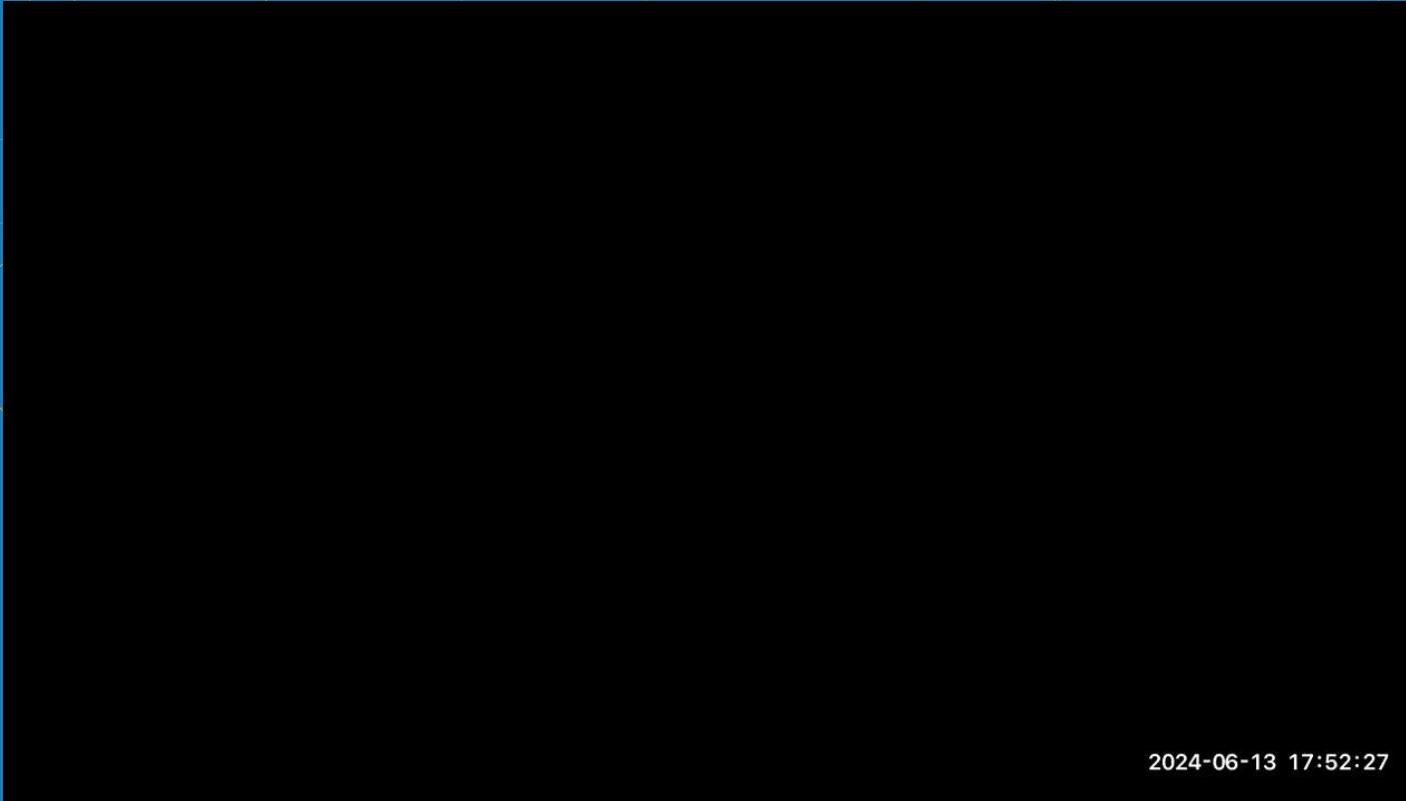
Video Generation, earlier this year



Sage Bionetworks

Source: OpenAI

Music Generation, earlier this year



2024-06-13 17:52:27



Sage Bionetworks

Types of Generative AI

Text Generation:

Creating coherent and contextually relevant text.
Example Models: GPT-4o, Llama 3, Claude Opus

Image Generation:

Creating new images from scratch or modifying existing images.
Example Models: DALL-E, Generative Adversarial Networks

Music Generation:

Composing original music or modifying existing tracks.
Example Models: ElevenLabs Music

Video Generation:

Creating new video content or animating existing images.
Examples: OpenAI Sora

Voice and Speech Generation:

Creating human-like speech and voice recordings.
Example Models: Meta Voicebox, ElevenLabs Multilingual v2

Code Generation:

Writing and generating programming code.
Example Models: Codex, GitHub Copilot.

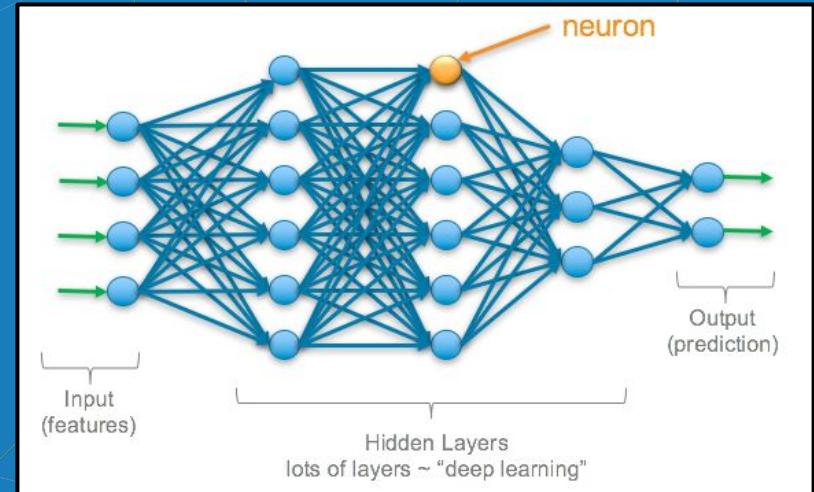


Large language models

LLMs are a type of generative AI that specializes in text.

Large Language Models (LLMs) are powerful computer programs that can read, write, and understand the rules of human language. LLMs read millions of sentences from books, articles, and websites. They learn patterns, grammar, and the meanings of words.

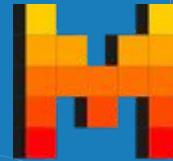
They use a type of machine learning called "deep learning."



Source: Stacy Ronaghahm



Large language (?) models

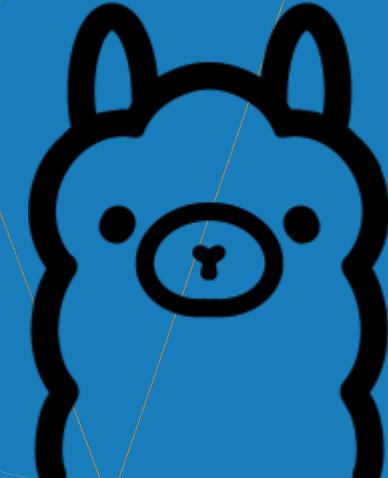


MISTRAL
AI_

 Claude 3



 Sage Bionetworks



Crash-course on concepts you might run across in LLMs

Zero-shot Learning: Model lets it do new tasks without any specific examples, using what it already knows.

Example: Model correctly describes a biological process that was never included in its training data.

Crash-course on concepts you might run across in LLMs

Zero-shot Learning: Model lets it do new tasks without any specific examples, using what it already knows.

Example: Model correctly describes a biological process that was never included in its training data.

Few-shot Learning: Few-shot learning teaches a model new tasks with just a few examples.

Example: Showing AI a few labeled pictures of cats and dogs so it can learn to tell the difference.



Crash-course on concepts you might run across in LLMs

Prompt Engineering: Prompt engineering helps you ask LLMs specific questions to get useful answers.

Example:

Writing "Can you write a short story about a futuristic brave knight?

vs

"Can you write a short story about a brave robot knight in 2120? The knight is named Sir ReSircher. In the story, they go on an important quest to defeat an autonomous AI that is bent on destroying all carbon-based life forms in the universe"

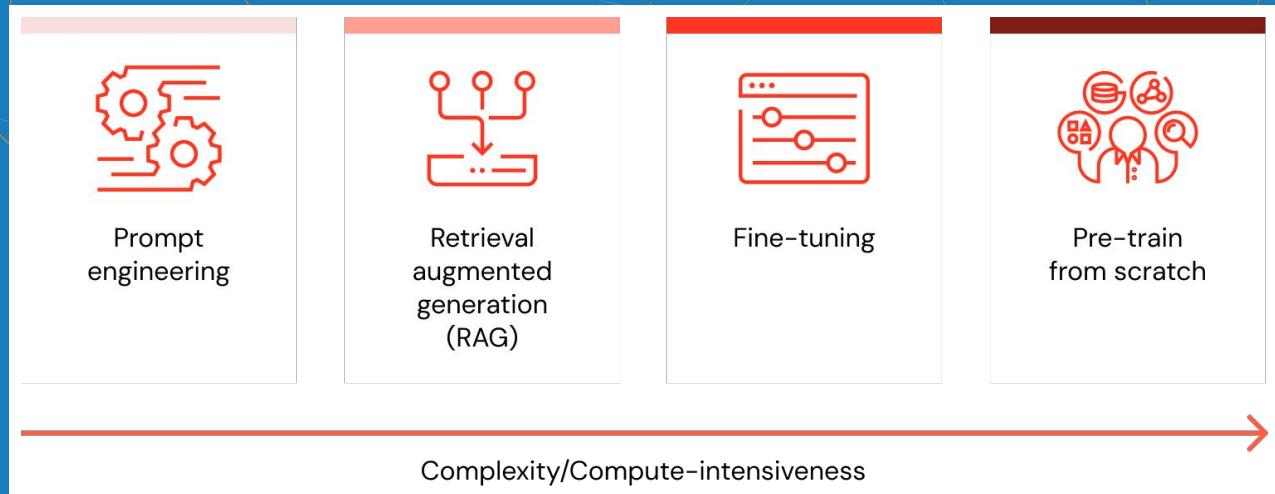
to get a story that better matches your criteria.

Crash-course on concepts you might run across in LLMs

Retrieval Augmented Generation (RAG):

Creation of text, or images, by an AI model that has been provided with additional relevant data.

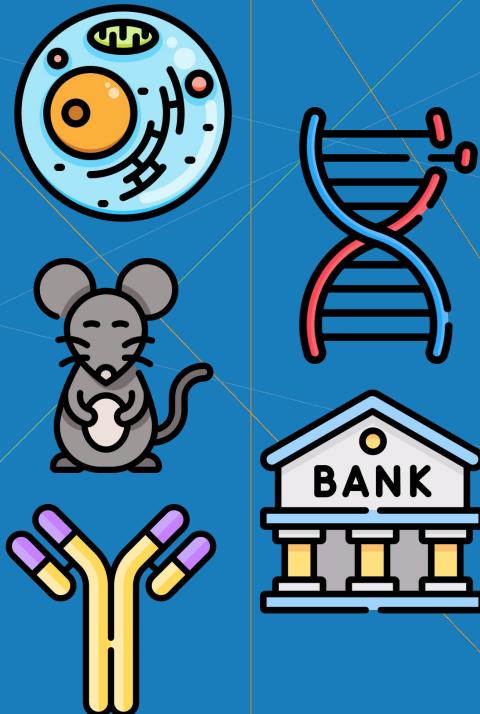
Example: LLMs don't have any information about what happened today. If you provide the text from today's newspaper, it will be able to answer questions about today's news.



Example use cases of how Sage has used LLMs

- Data Extraction and Formatting
- Adding Descriptive Metadata
- Summarizing Scientific Text
- Data Curation from Literature
- Text Simplification and Translation
- Code Development

Challenges in finding tools for research



- Reusable research tools, such as model organisms, cell lines, and antibodies, are essential in designing and executing successful biological experiments.
- **KEY CHALLENGE:** Reliable research is hard when reliable tools are not available. In NF, understanding:
 - which tools exist,
 - what they can be used for,
 - whether they are reliable is difficult

Challenges in finding tools for research

A variety of databases exist to help researchers find information about useful research tools:

The slide displays a collection of logos for biological databases and registries, arranged in a grid-like structure. The logos include:

- Cellosaurus: A logo featuring three overlapping circles in green, orange, and red.
- The Antibody Registry: A dark blue box containing the text "The Antibody Registry".
- addgene: A logo with a blue circular arrow icon followed by the word "addgene" in blue and grey.
- ATCC: A purple logo with a stylized 'A' and the text "ATCC".
- NIH Human Embryonic Stem Cell Registry: A logo featuring a small image of a stem cell and the text "NIH Human Embryonic Stem Cell Registry".
- RRID Portal: A logo featuring a magnifying glass icon and the text "RRID Portal".
- The Jackson Laboratory: A logo with a blue stylized 'J' and the text "The Jackson Laboratory".
- FlyBase: A hexagonal icon with a fly silhouette.
- MGD: A hexagonal icon with a mouse silhouette.
- RGD: A hexagonal icon with a rat silhouette.
- SGD: A hexagonal icon with a yeast cell silhouette.
- WormBase: A hexagonal icon with a worm silhouette.
- ZFIN: A hexagonal icon with a zebrafish silhouette.
- GOC: A hexagonal icon with a brain silhouette.
- A mouse icon: A small image of a brown mouse standing on its hind legs.

Challenges in finding tools for research

- Researchers have to visit multiple databases to find tools for a particular experiment
- It can be hard to keep track of changes in the central databases for research tools
- NF is a rare disease and has a specialized set of relevant tools; disease-agnostic databases that catalog **all** tools are overwhelming
- No place to deposit “unpublishable” *but valuable* observations about tool performance
- Researchers new to NF may have no prior knowledge of NF-related tools
- Existing tool databases generally do not have a space for pre-publication/in-development models
- **To address these challenges we built the NF Research Tools Database, prototyped using NF1-relevant tools**

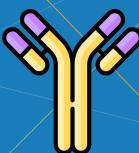
What type of research tools/resources are included?



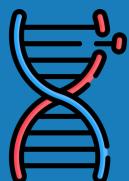
Cell lines: immortalized cell lines or share-able primary cell cultures



Animal models: Genetically modified models of any species as well as well-characterized and shareable xenograft models



Antibodies: Any commercially available or self-distributed antibody for NF-relevant targets



Genetic reagents: Plasmids, guide RNAs, useful primers, or other genetic tools for studying NF

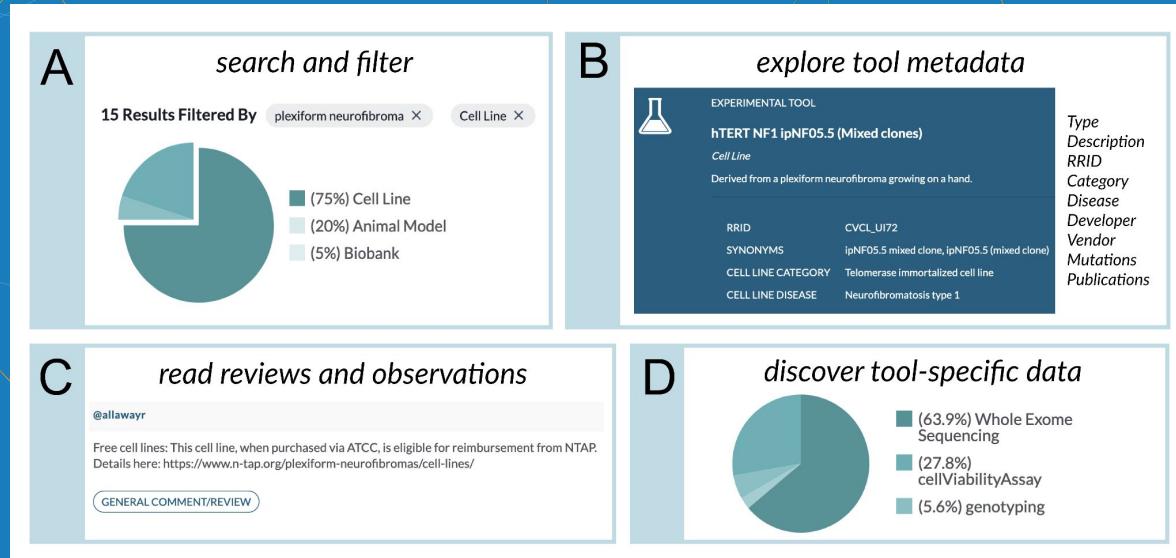


Biobanks: tissue or other biospecimen repositories with NF tumors or other NF-relevant biospecimens



Approach

- Built a user-friendly web application into the NF Data Portal to explore this database:
 - tools.nf.synapse.org



NF Research Tools Central

NF Research Tools Central aims to support the development of a robust research toolkit and lower the barrier of entry to neurofibromatosis (NF) research. The database includes NF-associated animal models, cell lines, antibodies, and genetic reagents and details on tool characteristics and sourcing, as well as observational and experimental data.

BROWSE TOOLS BY CATEGORY

Drill-down to explore specific types of NF research tools.

Animal Models

Antibodies

Genetic Reagents

Cell Lines

Biobanks

Features

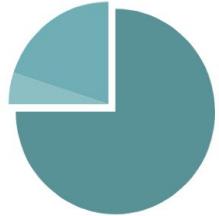
A

search and filter

15 Results Filtered By

plexiform neurofibroma X

Cell Line X



- (75%) Cell Line
- (20%) Animal Model
- (5%) Biobank

B

explore tool metadata



EXPERIMENTAL TOOL

hTERT NF1 ipNF05.5 (Mixed clones)

Cell Line

Derived from a plexiform neurofibroma growing on a hand.

RRID

CVCL_UI72

SYNONYMS

ipNF05.5 mixed clone, ipNF05.5 (mixed clone)

CELL LINE CATEGORY

Telomerase immortalized cell line

CELL LINE DISEASE

Neurofibromatosis type 1

Type
Description
RRID
Category
Disease
Developer
Vendor
Mutations
Publications

C

read reviews and observations

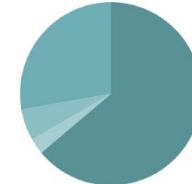
@allawayr

Free cell lines: This cell line, when purchased via ATCC, is eligible for reimbursement from NTAP.
Details here: <https://www.n-tap.org/plexiform-neurofibromas/cell-lines/>

GENERAL COMMENT/REVIEW

D

discover tool-specific data



- (63.9%) Whole Exome Sequencing
- (27.8%) cellViabilityAssay
- (5.6%) genotyping



Sage Bionetworks

Features

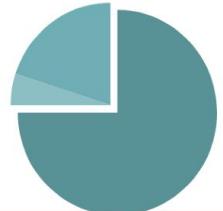
A

search and filter

15 Results Filtered By

plexiform neurofibroma X

Cell Line X



B

explore tool metadata



EXPERIMENTAL TOOL

hTERT NF1 ipNF05.5 (Mixed clones)

Cell Line

Derived from a plexiform neurofibroma growing on a hand.

RRID

CVCL_UI72

SYNONYMS

ipNF05.5 mixed clone, ipNF05.5 (mixed clone)

CELL LINE CATEGORY

Telomerase immortalized cell line

CELL LINE DISEASE

Neurofibromatosis type 1

Type
Description
RRID
Category
Disease
Developer
Vendor
Mutations
Publications

C

read reviews and observations

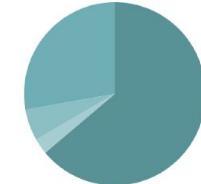
@allawayr

Free cell lines: This cell line, when purchased via ATCC, is eligible for reimbursement from NTAP.
Details here: <https://www.n-tap.org/plexiform-neurofibromas/cell-lines/>

GENERAL COMMENT/REVIEW

D

discover tool-specific data



Sage Bionetworks

Observations

“Observations” can be:

- natural history observations associated with a specific time point, such as ‘Nf1 Arg681*/Arg681* and Nf1 Δ4/Δ4 embryos failed to develop **beyond embryonic day 9.5**’
- general observations about a tool, such as ‘Optic gliomas in mice harboring the Nf1 Arg681* allele are more aggressive than those in mice with the Nf1 Δ31&32/Δ31&32 mutation.’

NATURAL HISTORY OBSERVATIONS

PRENATAL NEONATAL WEANLING JUVENILE ADOLESCENT ADULT



So, where now?

Lots of publications contain the types of “observations” that we want to collect.

- Researchers don’t have time to contribute all of this information, and it’s already (generally) published
- Prohibitively time-consuming to extract all of this info by hand from publications

So, where now?

Lots of publications contain the types of “observations” that we want to collect.

- Prohibitively resource-consuming to extract all of this info by hand
- Maybe LLMs could help automate this process?

Prompt Engineering

I am interested in learning more about the mouse models in this manuscript. Please extract a comprehensive list of temporal “observations” relevant to neurofibromatosis from this publication (<https://genesdev.cshlp.org/content/15/7/859.full.pdf>) that describe the natural history of the model as they relate to development or age. For example, an observation could be “The pigs developed tumor type X at Y months of age.” Do not include observations about humans with NF1. The observations should be formatted like the following JSON:

```
1 [  
2 {  
3   observationText: This is an example sentence.,  
4   observationType: [Body Length, Body weight, Coat Color, Di  
5   observationPhase: [prenatal, postnatal, null]  
6   observationTime: a double; the time during the development  
7   observationTimeUnits: [days, weeks, months, years]  
8 }  
9 ]
```

Typical Response

Based on the additional chunks retrieved from the paper, here are more temporal observations related to the development of mouse models with NF1:

```
1 v [
2 v {
3 v   "observationText": "NF1flox/- mice began to die at 10 mont
4 v   "observationType": ["Disease Susceptibility", "Mortality"
5 v   "observationPhase": ["postnatal"],
6 v   "observationTime": 10,
7 v   "observationTimeUnits": "months"
8 v },
9 v {
10 v     "observationText": "NF1\ud83c\udc00-embryos are characterized
11 v     "observationType": ["Organ Development", "Disease Suscept
12 v     "observationPhase": ["prenatal"]
13 v     "observationTime": 13.5,
14 v     "observationTimeUnits": "days"
15 v   }
16 v ]
```

These observations provide specific time points at which certain phenotypes are observed in the mouse models. The first observation indicates a survival rate issue starting at 10 months of age in NF1^{flox/-} mice, while the second observation details the phenotype of embryonic lethality at 13.5 days of embryonic development in NF1Δ/- embryos.

Would you like to continue with the retrieval of more content from the paper for further observations?

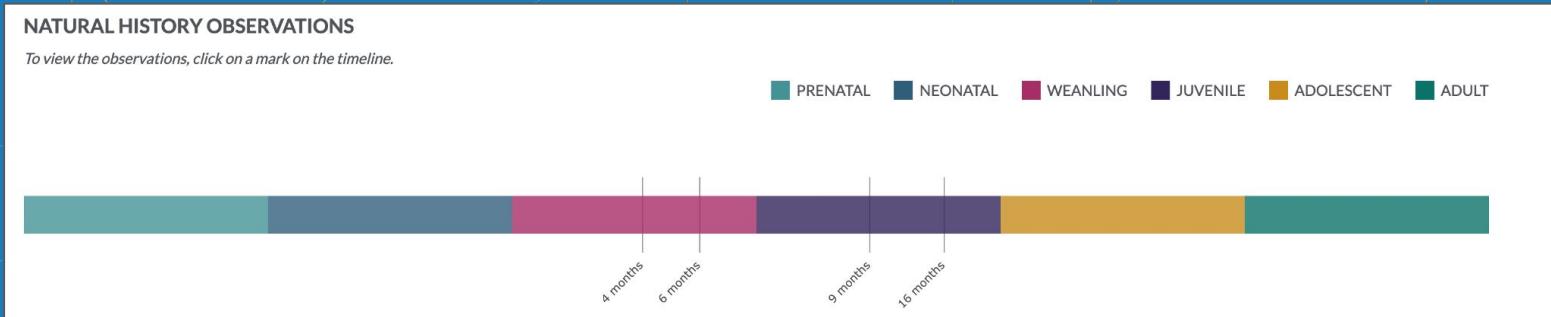
Results/Validation

- ChatGPT extracts:
 - the type of observations we are looking for
 - correctly assigns the phase (“prenatal” or “postnatal”)
 - a type of observation (though do note that it’s suggesting a type – mortality – that is not in our list of terms – it is very important to validate these results against our schema),
 - the time, and units of the time of the observation.
- Looking at the paper for confirmation, we can see that both of these observations are correct.

flox mice that exhibited a typical wild-type survival profile, NF1^{flox/-} mice began to die at 10 mo of age, and only 80% survived beyond 14 mo (Fig. 1E). Thus, the

NF1^{Δ/-} embryos are indistinguishable from NF1^{-/-} embryos, which are characterized by an enlarged heart, peripheral edema, hemorrhage, small unpigmented eyes, and which die at E13.5 because of cardiac abnormalities

Add it to the portal



🕒 Al-extracted

🕒 4 months

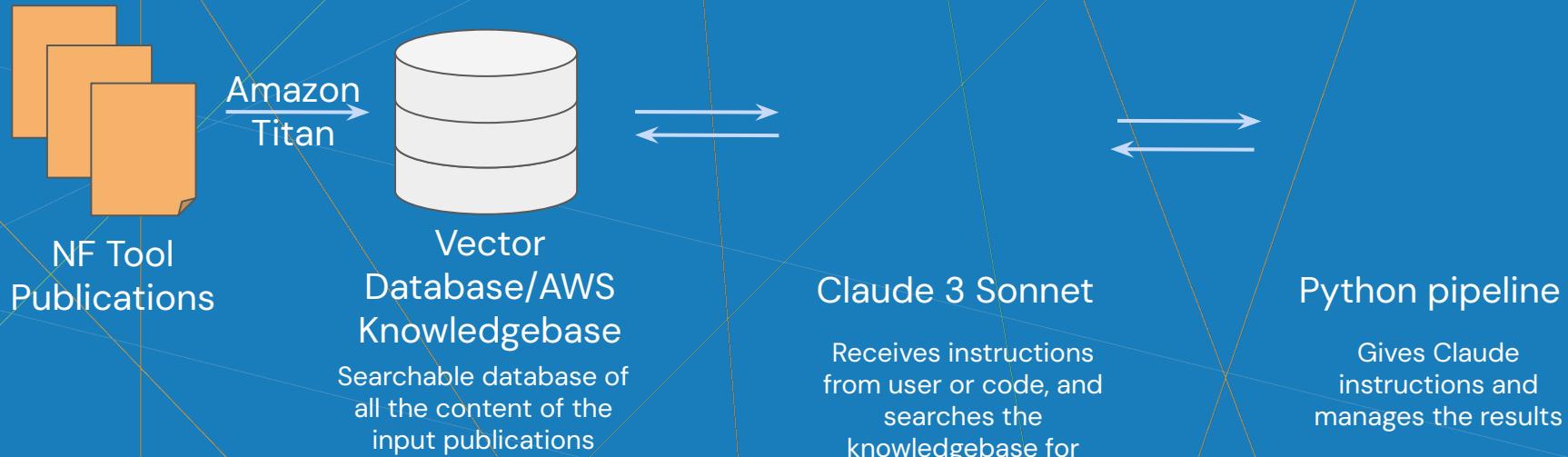
The NF1 minipig closely mimics the age of onset for NF1-related phenotypes in humans. CALMs are often present at birth in humans, and neurofibromas tend to develop around puberty. In the NF1 minipig model, CALMs were observed perinatally in all pigs, and minipig neurofibromas develop at around 4 months of age, with puberty occurring around 5-6 months.

Reference

DISEASE SUSCEPTIBILITY

Where are we today and where do we want to go?

- Have built a proof of concept semi-automated pipeline to begin to scale up to the entire tools database



- Currently have around 800 AI-extracted observations about animal models and cell lines in NF Tools Central
 - Extracted from approximately 400 open-access publications
 - Want to extract more, but it's tricky to balance false positive and false negative rates
- We could use much of this technology to build an NF Tools chatbot that allows anyone to ask questions about these publications and the tools described or used in them



Making science more accessible

Example:

"The NF1 phenotype is clinically variable; however, a few variants have been found to be compatible with partial protein production and correlate with specific phenotypes. These distinct phenotypes involve either the nerve sheath (Schwann cells (SC)) with minimal melanocytic involvement or melanocytes with no nerve sheath involvement. A severe "spinal NF" phenotype is associated with G848R and R1276Q and is characterized by a massive internal tumor burden, with neurofibromas (originating from SC) at each spinal nerve root and extreme enlargement of most peripheral nerves. These individuals are at great risk of spinal cord compression, pain, and malignant change but, have a mild pigmentary phenotype. In contrast, relatively mild phenotypes are associated with delM992 and R1809C which are characterized by CALMs (originating from melanocytes) and no neurofibromas (no SC involvement) or risk for malignancy. Hence, distinct cell populations are affected by each mutation, resulting in different clinical phenotypes with different implications for later development of malignancy. Our overall goal is to determine how these genotypes affect the function of neurofibromin and its interactions with other proteins. We will test the hypothesis that neurofibromin differentially interacts with binding partners in a cell type-specific manner, and that mutations differentially disrupt those interactions. The proposed research will lead to identification of new neurofibromin functions and associated therapeutic targets. Clinical Impact. This proposal directly addresses the role of NF1 mutations in cancer. We compare mutations that are and are not associated with malignancy in two different clinically relevant cell types. Identifying binding partners and determining those that are associated with malignancy may provide new therapeutic targets for cancers driven by NF1 mutations."

Plain language summaries (PLSs)

Prompt: Please modify this abstract so that it is at a 6th-grade reading level.

"The NF1 phenotype is clinically variable" →
"NF1 is a health condition that can look different in different people"

"We will test the hypothesis that neurofibromin differentially interacts with binding partners in a cell type-specific manner, and that mutations differentially disrupt those interactions." → **"We think that neurofibromin interacts differently with other proteins in different cells, and that the gene changes can change those connections".**



NF1 is a health condition that can look different in different people. There are some versions of this condition that can make only part of a protein, which leads to specific symptoms. These symptoms can affect either the nerves in our body (called Schwann cells) with little effect on skin color cells (melanocytes) or the other way around.



Some people with NF1 have a severe type called "spinal NF," linked to two specific gene changes (G848R and R1276Q). They have many tumors inside their body, especially in the nerves of the spine and other big nerves. This can lead to serious problems, like spinal cord compression, pain, and even cancer. However, these people usually have only mild changes in their skin color.

On the other hand, some people have milder forms of NF1, connected to two other gene changes (delM992 and R1809C). These people have spots on their skin (from melanocytes) but no tumors in their nerves (no Schwann cell involvement) and no risk for cancer.

The goal of our study is to learn how these gene changes affect the way the protein (neurofibromin) works and how it connects with other proteins. We think that neurofibromin interacts differently with other proteins in different cells, and the gene changes can change those connections. Our research will help us learn more about what neurofibromin does and find new ways to treat the problems caused by NF1.

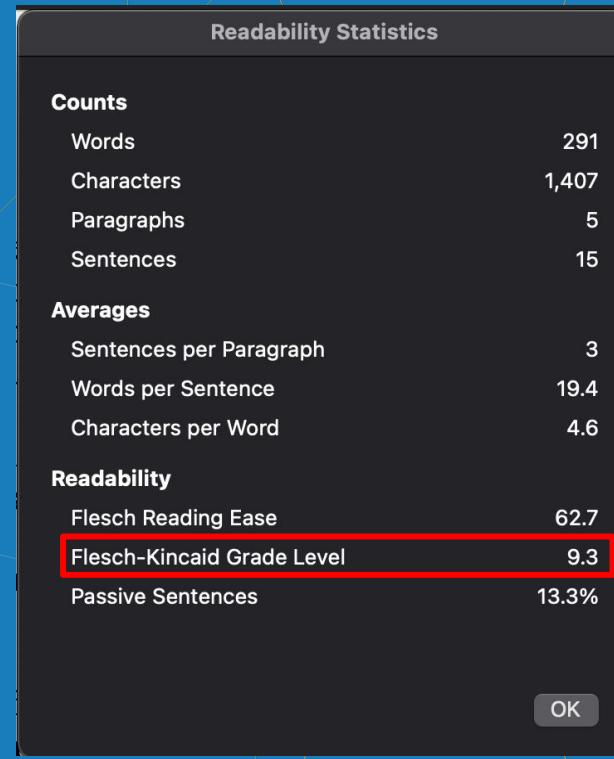
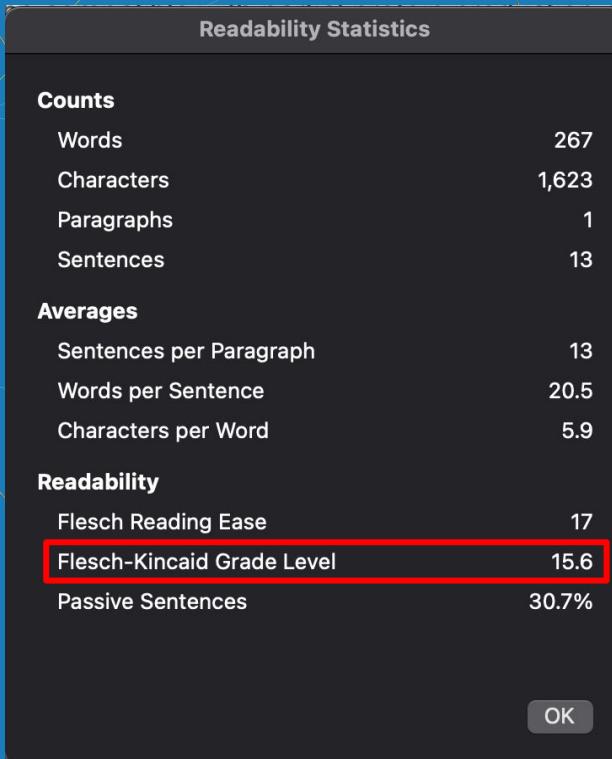
Our study is important because it focuses on how NF1 gene changes can lead to cancer. We will compare gene changes that are and are not connected to cancer in two different types of cells. By figuring out which connections are linked to cancer, we may find new ways to treat cancers caused by NF1 gene changes.



Sage Bionetworks

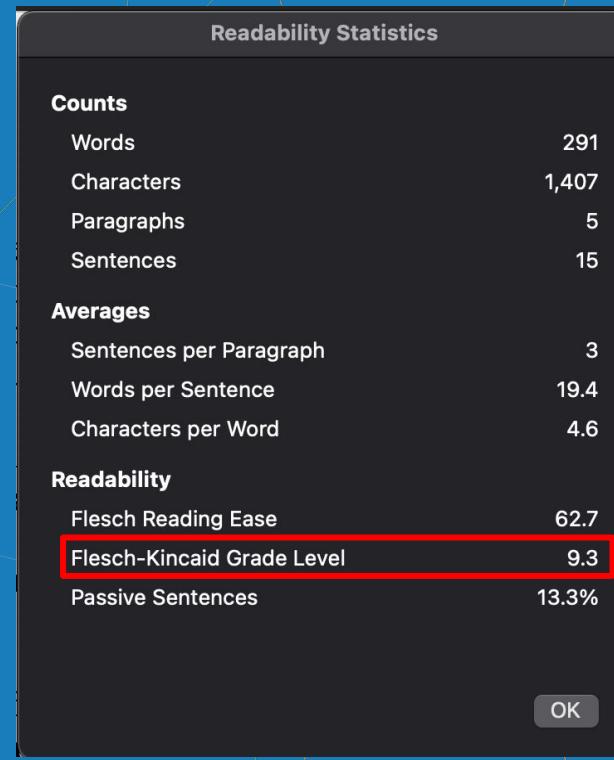
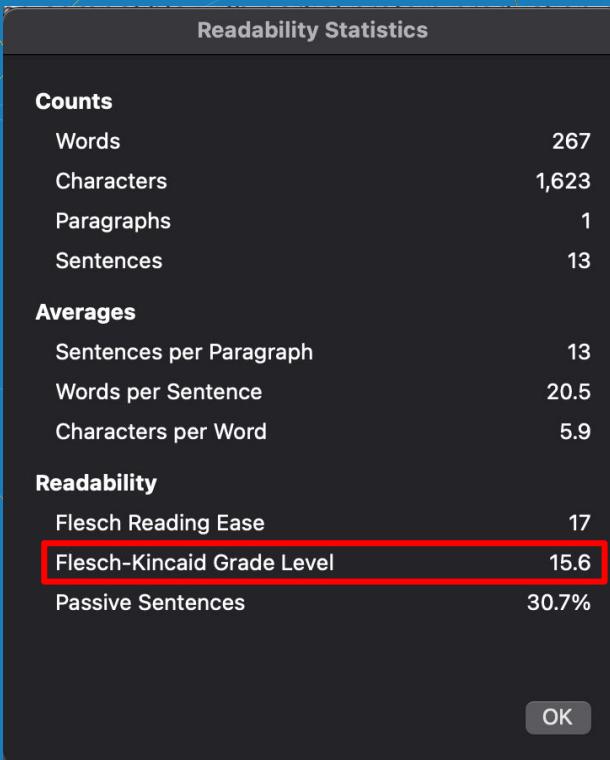
Layperson-friendly translation

Prompt: Please modify this abstract so that it is at a 6th-grade reading level.

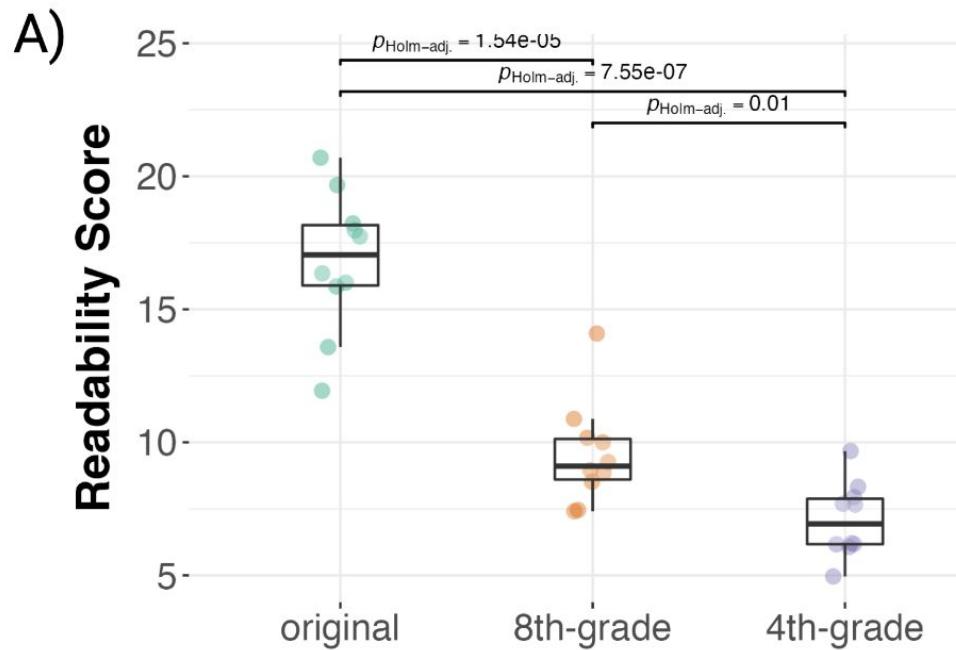


Layperson-friendly translation

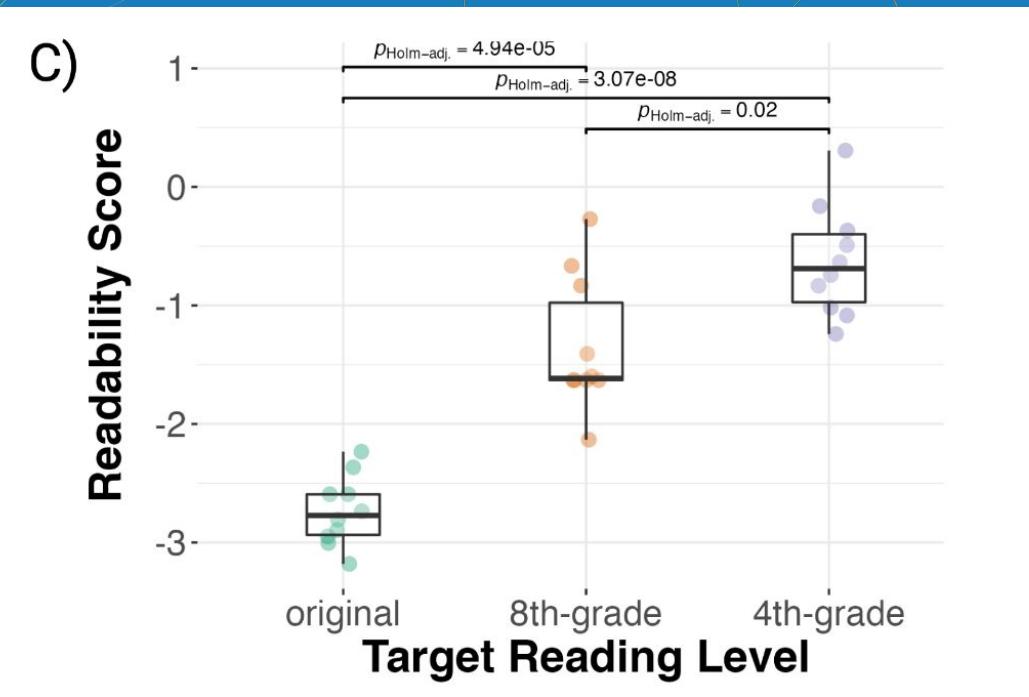
Prompt: Please modify this abstract so that it is at a 6th-grade reading level.



Preliminary Data - Flesch-Kincaid Grade Level



Preliminary Data – CommonLevel (deep learning-based)



Time to get some hands on experience!

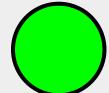
- We hope these examples inspire you to think about some ways to apply LLMs in your daily work
- We've developed a series of hands-on tasks that you can try to get some experience using large language models

Working Session Tasks:

Let us know if you need help by flipping your card



Help!



Good

We have prepared three workshop tasks. Select one, two or all three!

- Due to free account limits, we suggest doing the tasks in the order of your interest in them (most interested first).
- These tasks were all tested with the GPT-4o model, but they should generally work with all current LLMs like Claude Sonnet, Llama 3b
- Remember – LLMs sometimes do unpredictable things, but you are in the driver's seat. We're happy to help with prompt suggestions!
- This would be a good time to log in to ChatGPT, Claude, Meta.ai or another LLM.



Sage Bionetworks

Working Session Tasks:

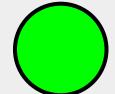
Exercise 1: Summarize a large block of text from a publication and extract the information that is useful to you.

Learning Outcome: Participants will learn how to use ChatGPT to effectively summarize lengthy text, focusing on extracting and retaining the most relevant and useful information.

Extra Challenge: Practice summarizing the text for different audiences.

Exercise 1 link [here!](#)

Let us know if you need help by flipping your card



Working Session Tasks:

Let us know if you need help by flipping your card



Help!



Good

Exercise 2: Extract data from given publications. Learn how to plug in an image and extract the information within it.

Learning Outcome: Participants will understand how to use ChatGPT to extract and interpret data from images in scientific publications.

Extra Challenge: Develop prompts to extract detailed insights from complex figures and tables!

Exercise 2 link [here!](#)



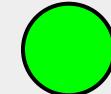
Sage Bionetworks

Working Session Tasks:

Let us know if you need help by flipping your card



Help!



Good

Exercise 3: Use the provided example dataset and a metadata template. Try to fill in the annotations using ChatGPT.

Learning Outcome: Participants will learn how to use ChatGPT to automate the process of annotating metadata, improving efficiency and accuracy in data management without coding.

Extra Challenge: Try using ChatGPT to double check itself!

Exercise 3 link [here!](#)



Sage Bionetworks

Working Session Tasks:

Let us know if you need help by flipping your card



Exercise 1: Summarize a large block of text from a publication and extract the information that is useful to you.

Learning Outcome: Participants will learn how to use ChatGPT to effectively summarize lengthy text, focusing on extracting and retaining the most relevant and useful information.

Extra Challenge: Practice summarizing the text for different audiences.

Exercise 1 link [here!](#)

Exercise 2: Extract data from given publications. Learn how to plug in an image and extract the information within it.

Learning Outcome: Participants will understand how to use ChatGPT to extract and interpret data from images in scientific publications.

Extra Challenge: Develop prompts to extract detailed insights from complex figures and tables!

Exercise 2 link [here!](#)

Exercise 3: Use the provided example dataset and a metadata template. Try to fill in the annotations using ChatGPT.

Learning Outcome: Participants will learn how to use ChatGPT to automate the process of annotating metadata, improving efficiency and accuracy in data management without coding.

Extra Challenge: Try using ChatGPT to double check itself!

Exercise 3 link [here!](#)



Wrap-up

- Have any questions about LLMs or AI?
Let's talk!
- Have ideas for how we should integrate
AI into the NF Data Portal? Let us know!
- Want to contribute to the NF Tools
Central? Share your author-manuscripts
with us that describe how you have
developed or used NF tools!
- Please fill out our post workshop survey
(super quick!) using the QR code on the
right →

